

( 続紙 1 )

京都大学	博士 ( 情報 学 )	氏名	葉 呈陽 (Chengyang Ye)
論文題目	A Study on Multi-Granularity Representation Learning of Time Series Data (時系列データのマルチグラニューラリティ表現学習に関する研究)		
(論文内容の要旨)			
<p>Time series data mining is critical in various domains, notably healthcare and finance, where its applications and implications are particularly significant. This field is driven by representation learning, a key technology that transforms unstructured time series data into organized, meaningful vectors. This transformation is vital for effective data analysis, pattern recognition, and strategic decision-making. Traditional methods in this field have often borrowed techniques from computer vision (CV) and natural language processing (NLP). While these approaches can be effective, they may need to fully encapsulate the unique correlations and sequential patterns intrinsic to time series data. Furthermore, an essential aspect often overlooked is the multi-granularity nature of time series data. Understanding interactions and dependencies across different levels of granularity can unveil a more nuanced and comprehensive representation, leading to more precise and reliable predictions and decisions.</p> <p>This research addresses these challenges by investigating representation learning models of time-series data that operate across various granularities, a critical element for capturing the intricate dynamics of temporal data at different detail levels. The study encompasses four topics in multi-granularity representation learning of time series data. The research demonstrates the enhanced performance of the proposed methods in downstream tasks such as classification, prediction, and similar data retrieval, compared to existing models, by conducting experiments with public datasets.</p> <p>The first topic investigates the timestamp level, delving into the fine-grained details of the data. This level of granularity is essential for high-resolution insights, capturing subtle yet significant patterns and minor fluctuations over time. Such fine-grained analysis is crucial for sensitive applications where minute details can substantially impact.</p> <p>In contrast, the second topic shifts focus to a broader, more macroscopic view at the segment level. This perspective aims to capture and understand overall trends and shifts occurring over extended time intervals, providing a coarse-grained representation of the time series. This type of analysis is beneficial in scenarios where long-term trends and overarching patterns are the primary interest.</p> <p>A novel cross-granularity representation model is introduced to bridge these distinct levels of detail. This innovative model seamlessly integrates both fine-grained and coarse-grained representations. Doing so offers a more</p>			

holistic and comprehensive understanding of time series data. Such an integrated approach is especially beneficial in complex analytical scenarios requiring insights into immediate, detailed patterns and longer-term, broader trends.

Additionally, the research explores the extension of this model to the realm of streaming time series data, which presents its unique challenges. Streaming data, characterized by its continuous influx and real-time nature, demands prompt processing and the capacity to adapt to rapidly evolving patterns. Demonstrating the model's effectiveness in this dynamic environment underscores its versatility and potential for widespread application across various industries, particularly those where real-time data analysis is paramount.

In conclusion, this thesis significantly advances time series data representation learning across multiple granularities. The proposed models enhance our understanding of temporal data and open new avenues for research and application in critical sectors like healthcare and finance.

(論文審査の結果の要旨)

時系列データの分析・管理は、特にヘルスケアや金融などの様々な分野で重要な役割を担っており、その応用と影響は特に顕著である。構造化されていない時系列データを効率的に分析・管理するためには、データの様々な粒度を考慮して意味のあるベクトル表現を獲得する表現学習が不可欠である。

本論文では、時系列データの多粒度表現学習における四つの重要な技術的課題に取り組んだ成果をまとめている。(1) タイムスタンプレベルの表現学習、(2) セグメントレベルの表現学習、(3) 粒度横断の表現学習、そして(4) ストリームデータへの適用である。具体的には、これらの課題に以下の成果を挙げている。

第一に、タイムスタンプレベルの表現学習では細粒度の表現を学習する手法を提案した。提案手法は、時間の経過に伴う微妙なパターンや小さな変動を捉えることに特化しており、高解像度の洞察が求められる敏感なアプリケーションにおいて重要である。表現性能を向上させるため、ローカルバイナリパターン(LBP)手法を拡張して多次元時系列データに適用可能にした。また、それを効率的に処理する自己注意機構を設計した。新しい教師なし学習アプローチを導入し、学習効率の向上も図った。分類と回帰を下流タスクとした実験により、提案手法の細粒度特徴に対する効果を確認した。

第二に、セグメントレベルの表現学習では、時系列データの部分系列特徴を捉えるための教師なし学習モデルを提案して、より広い時間間隔にわたる広範囲なトレンドや変動を理解して取り込むことで時系列データの粗粒度の表現を学習する。提案手法では、ガウス過程によって計算された共分散を自己注意機構に導入して、部分系列の関連性を捉える。検索を下流タスクとした実験により、提案手法が時系列データの粗粒度表現において効果的であることを確認した。

第三に、異なる粒度の表現学習モデルを統合するために、粒度横断の表現学習モデルを提案している。このモデルは、細粒度と粗粒度の表現を統合し、それぞれの強みを活かして時系列データのより包括的な表現を獲得できる。分類と回帰を下流タスクとした実験により、複数の公開データセットを用いてその効果を確認した。

第四に、ストリーミング型データにも適用可能なセグメントレベルの学習手法を提案している。これにより本論文で提案する表現学習がオンライン処理にも拡張可能であることを確認した。

本論文は、時系列データを効率よく管理・分析するための多粒度表現学習における四つの重要な課題に対する解決手法をまとめたもので、学術上、および、實際上、寄与するところが少なくない。よって、本論文は博士(情報学)の学位論文として価値あるものと認める。また、令和5年12月13日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。また、本論文のインターネットでの全文公表についても支障がないことを確認した。